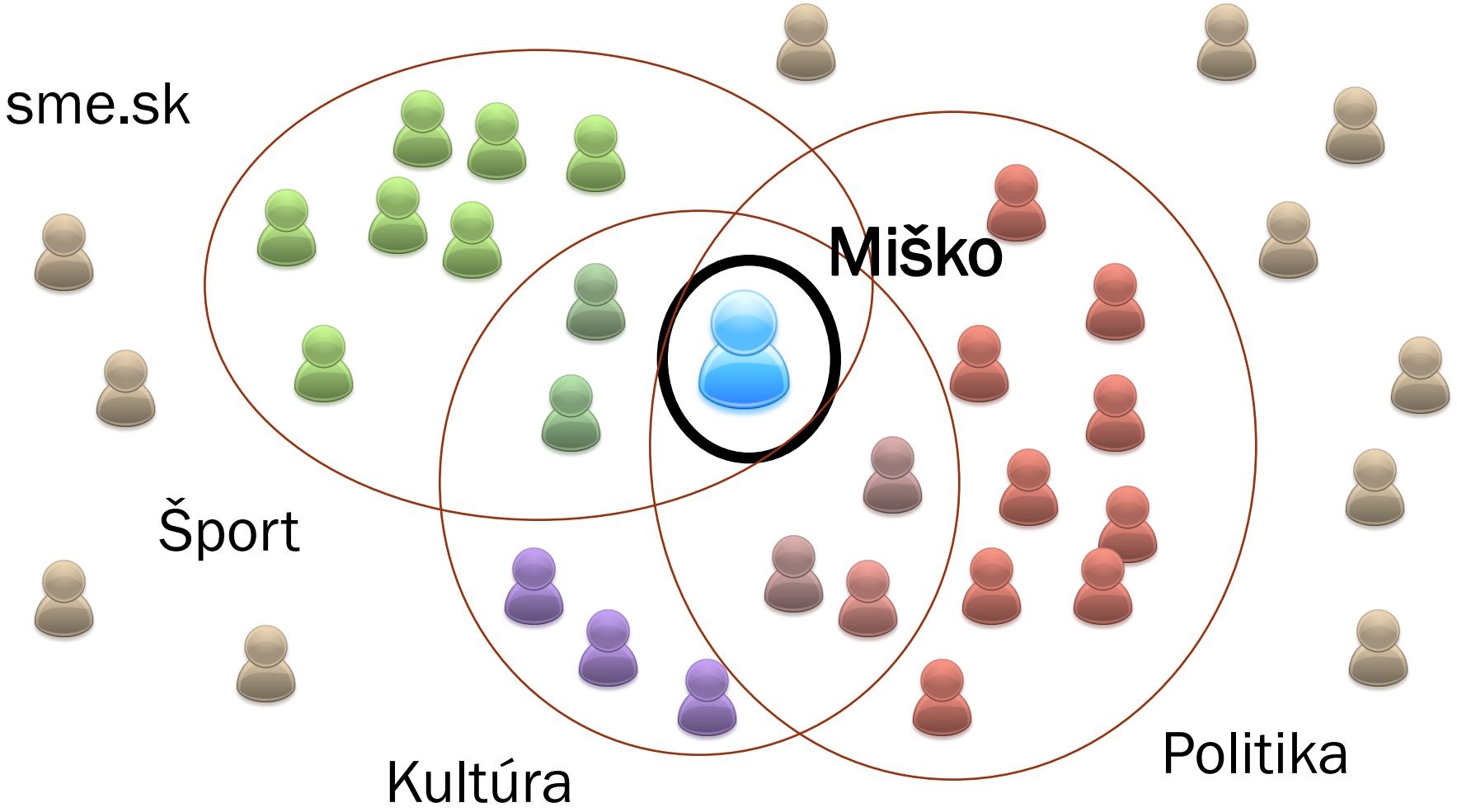
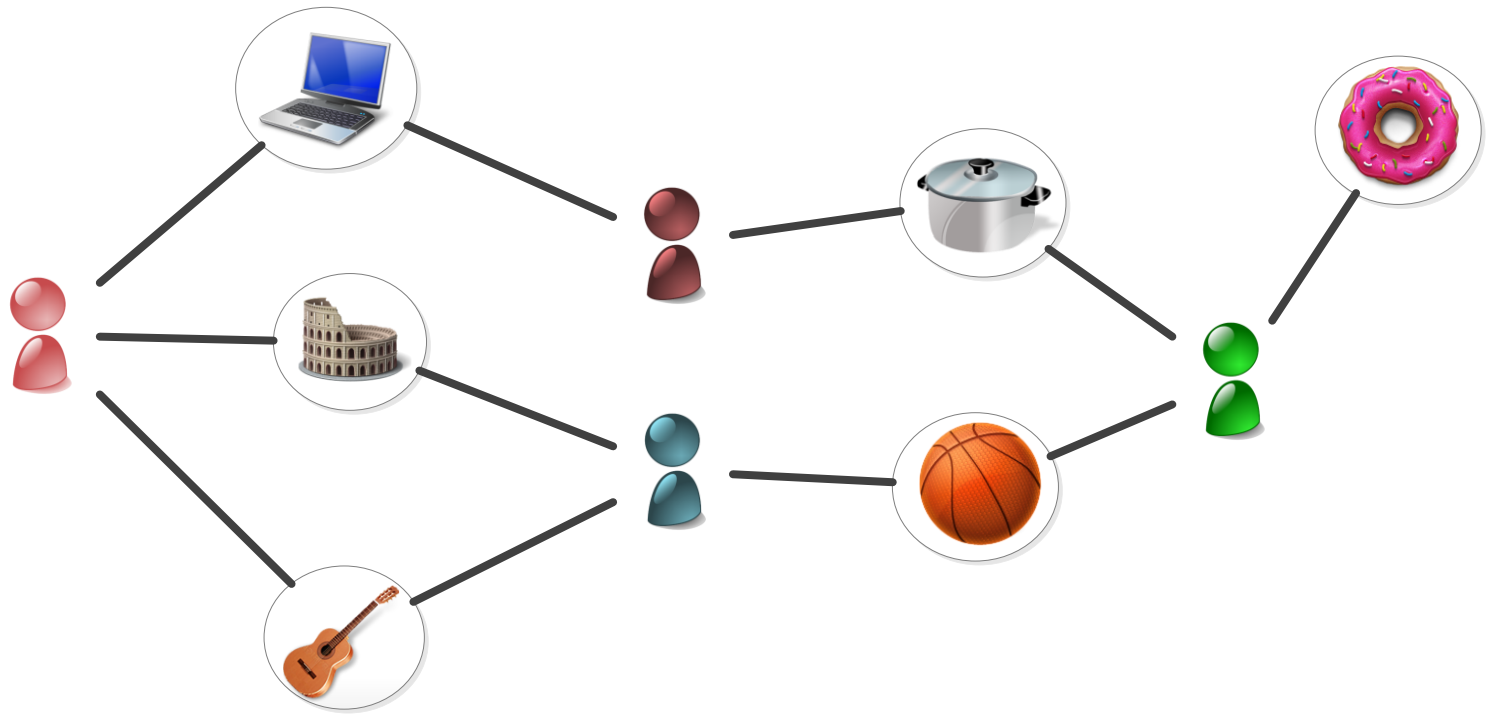
Identifikácia virtuálnych komunít v otvorenom priestore na webe

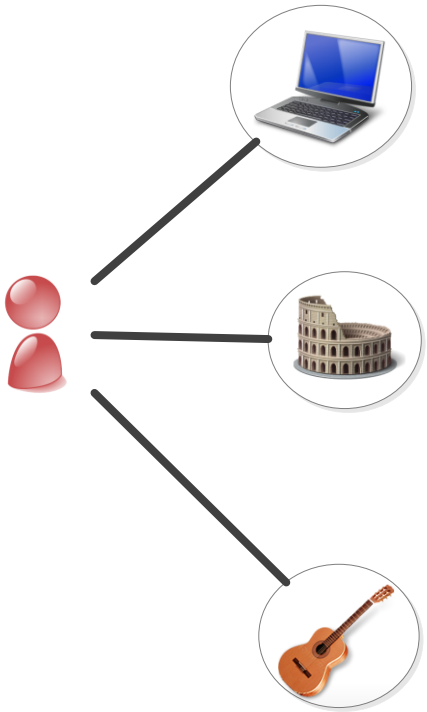
Marián Hönsch

Vedúci projektu: Michal Barla,PhD.

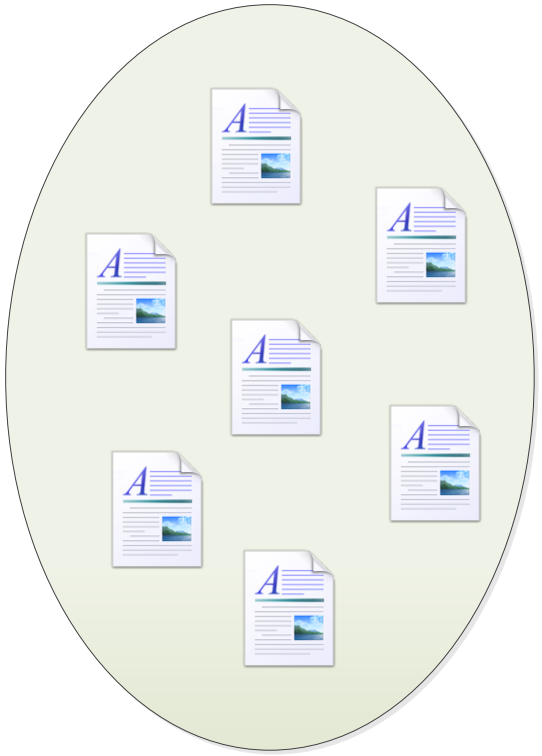
Spojenie ľudí s rovnakými záujmami (ale nie všetkými)



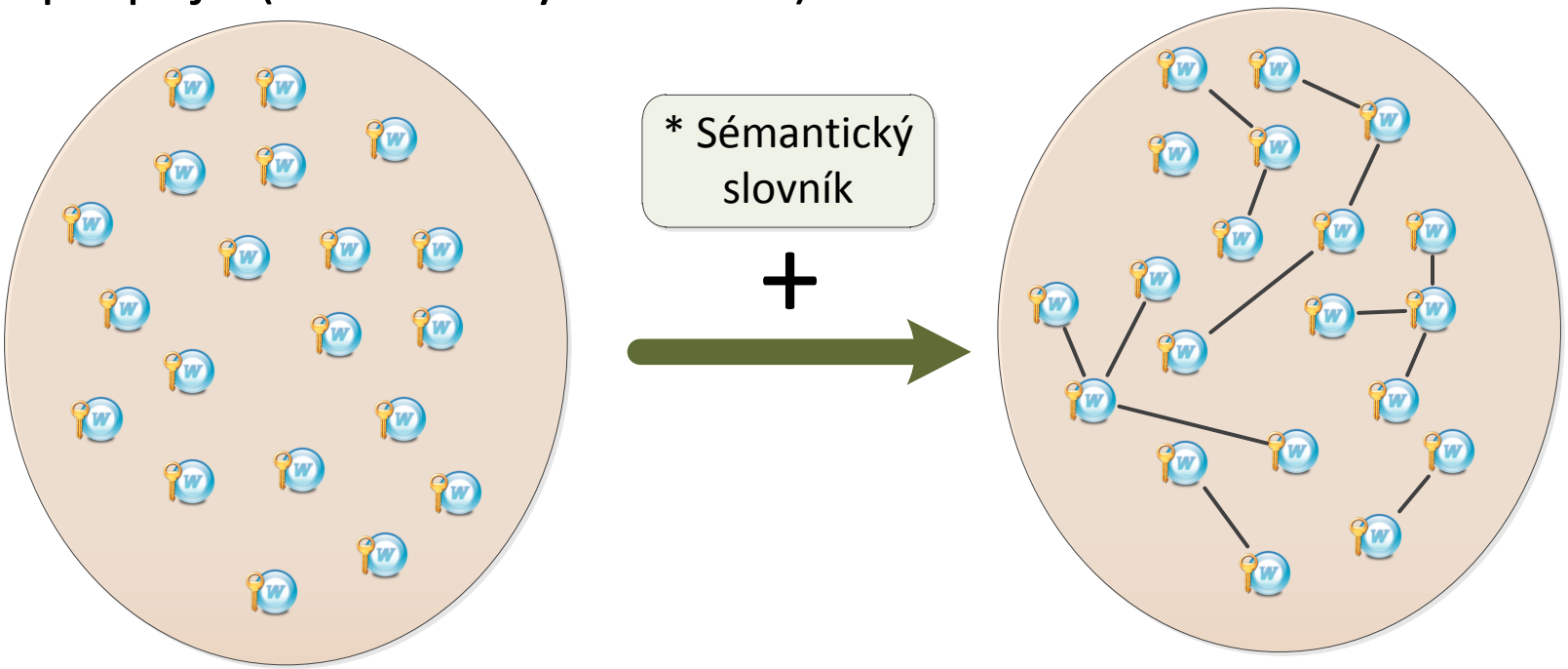




Záujmy sú odvodené od článkov, ktoré prečítal



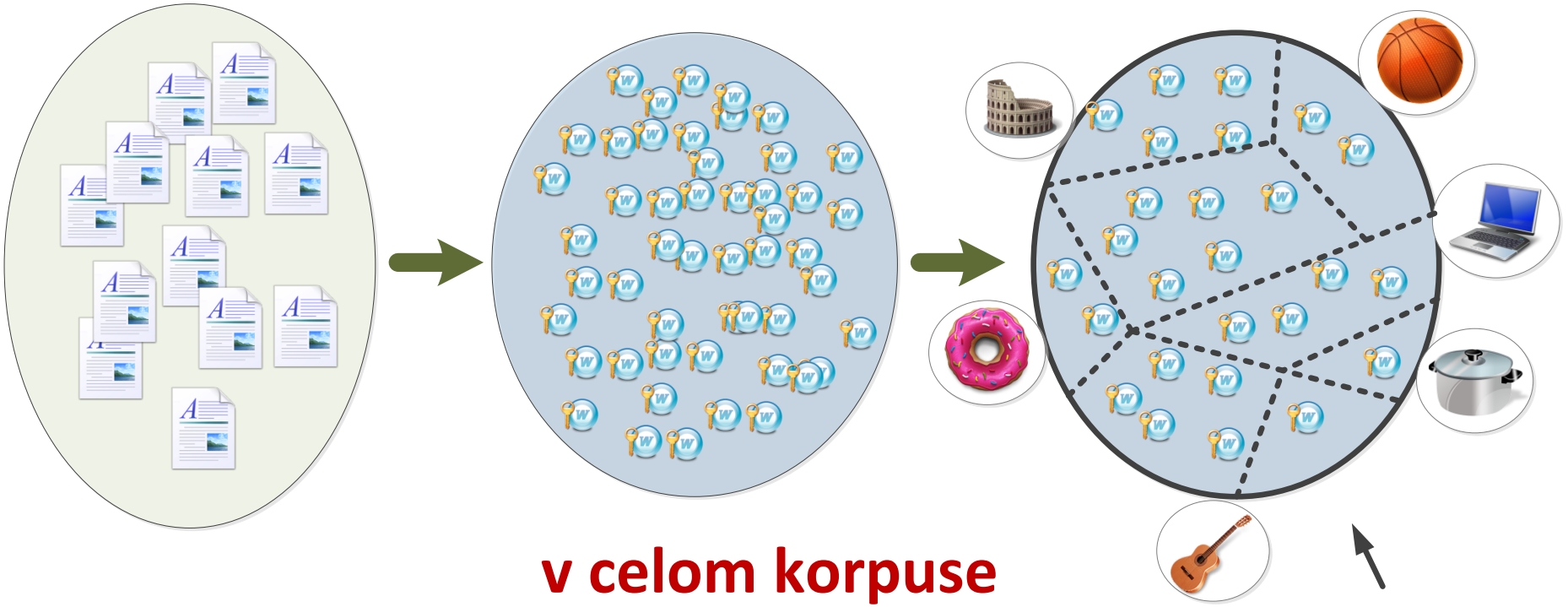
Na aplikovanie algoritmov pre hľadanie virtuálnych skupín musíme slová návzájom prepojiť (sémanticky obohatiť)



Všetky články

Všetky slová

Všetky záujmy



v celom korpuse

globálne (celkové) záujmy

Overenie 1.

Porovnávanie článkov za pomoci vlastného sem. slovníka

Pokrytie:	0.779
Presnosť:	0.951
f ukazovateľ:	0.856

Ukážka 1.

1. skupina, 15 slov
blackburn newborn chromosome cleavage mutate gnome enzyme discovery hereditary mammal disposition genetic mitochondria molecule syndrome

2. skupina, 26 slov
orava tvrdosn habovka dolny kubin roofer potok zuberec babin namestovo merry vessel zazriva trstena polhora poruba sihelne namestovsk eho hrustn rabca istebne liesek medzibrod benadovo jablonka klin

3. skupina, 47 slov
pianist orchestra songwriter studio legend bass trio repertoire guitarist dramaturgy pop percussionist cult improvise rhythm sting
blues debut gig bassist drummer compose guitar album genre composite lipa musician soul keyboardist saxophone legendary tireless
singer roll swing percuss rock saxophonist quartet headline jazz
sing funk virtuoso harmonica drum

Overenie 2.

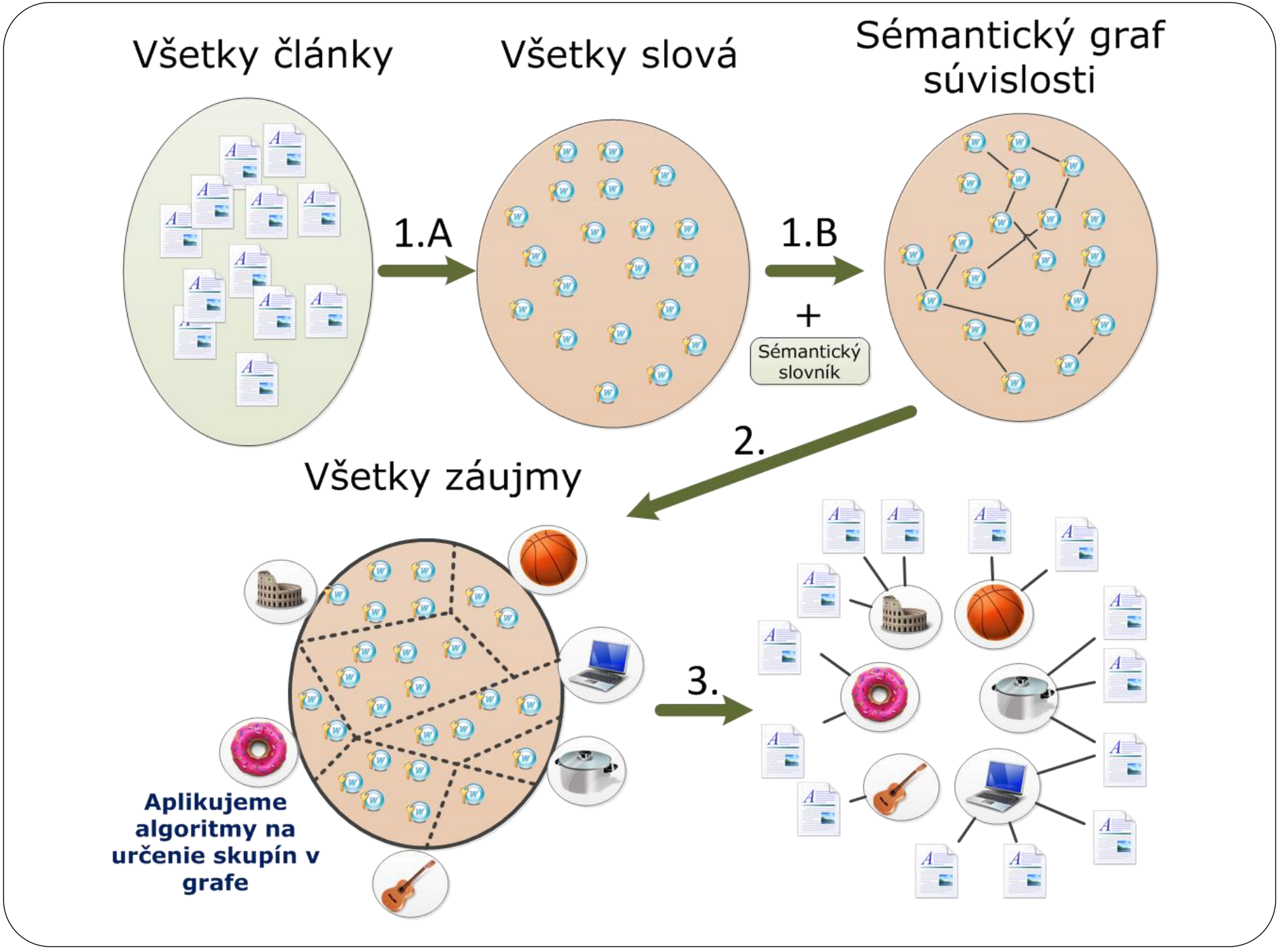
Odporúčanie a predikcia

Odstránených N najčít. článkov	Naša metóda		Náhodné skupiny	
	Presnosť	Podmn. článkov	Presnosť	Podmn. článkov
0	31,2%	19,6%	27,7%	10%
10	27,6%	35%	25%	12,4%
20	19,9%	40%	15,9%	16,5%
30	17,8%	42%	12%	20,7%
40	15 %	45%	9,5%	21%

Overenie 3.

Odporúčanie a predikcia inak

Odstránených N najčít. článkov	Naša metóda	
	Presnost'	Podmn. článkov
0	23,5%	37%
10	26%	39%
20	21%	40%
30	22%	39%
40	20 %	41%


Všetky články
Všetky slová
Sémantický graf súvislosti
1.A
1.B
+
Sémantický slovník
2.
Všetky záujmy
3.
Aplikujeme algoritmy na určenie skupín v grafe

Overenie 4.

Rozdelenie článkov podľa záujmov ?=
rozdelenie podľa sekcia + kategória

	TOP 1	TOP 2	TOP .3	ALL
Podiel zhlukov s 1 dom. kombináciou	48,50%	50,11%	47,65%	26,33%
Podiel zhlukov s 2 dom. kombináciami	20,89%	18,12%	15,07%	9,66%
Podiel zhlukov s 3 dom. kombináciami	4,85%	2,59%	3,68%	10,98%
Podiel zhlukov s 4 dom. kombináciami	1,19%	0,00%	0,00%	12,87%
(zhluky s dom. kom.) / (všetky zhluky)	75,43%	70,82%	66,39%	59,84%
Priemerný počet kombinácií na zhluk	2,46	3,68	4,56	5,38
Priemerný počet zhlukov na komb.	4,01	7,15	9,46	11,36

Overenie 4.

Najpodobnejšie sekcie a kategórie

1.	Korzár Košický a SME Online Z domova	6.	Korzár Košický a Korzár Prešovský
2.	SME Online Zahraničie a SME Online Z domova	7.	SME Online Z domova a SME Print Spravodajstvo
3.	Primár Choroby a Primár Bolesť	8.	Korzár Košický a Bratislava Bratislava
4.	Bratislava Bratislava a SME Online Z domova	9.	Korzár Košický a Korzár Tatranský
5.	Počítače Správy a Počítače Téma	10.	Primár Choroby a Primár Tehotenstvo